

UTILITY		Attorney Docket No.	113122	Total Pages	30
PATENT APPLICATION TRANSMITTAL		First Named Inventor or Application Identifier			
		Yuguang Wu			
For new nonprovisional applications under 37 CFR 1.53(b)		Express Mail Label No.	EL580483135US		
APPLICATION ELEMENTS See MPEP chapter 600 concerning utility patent application contents.		ADDRESS TO:	Assistant Commissioner for Patents Box Patent Application Washington, D.C. 20231		
<div>1. <input checked="" type="checkbox"/> Fee Transmittal Form (submit an original, and a duplicate for fee processing)</div> <div>2. <input checked="" type="checkbox"/> Specification [Total Pages 14] (preferred arrangement set forth below)<ul style="list-style-type: none">- Descriptive title of invention- Cross References to Related Applications- Statement Regarding Fed sponsored R&D- Reference to Microfiche Appendix- Background of the Invention- Brief Summary of the Invention- Brief Description of the Drawings (if filed)- Detailed Description- Claim(s)- Abstract of the Disclosure</div> <div>3. <input checked="" type="checkbox"/> Drawing(s) (35 USC 113) [Total Sheets 5]</div> <div>4. Oath or Declaration [Total Pages 4]<ul style="list-style-type: none">a. <input checked="" type="checkbox"/> Newly executed (original or copy)b. <input type="checkbox"/> Copy from a prior application (37 CFR 1.63(d)) (for continuation/divisional with Box 15 completed) [Note Box 15 below]i. <input type="checkbox"/> DELETION OF INVENTOR(S) Signed statement attached deleting inventor(s) named in the prior application, see 37 CFR 1.63(d)(2) and 1.33(b)</div>		<div>5. <input type="checkbox"/> Microfiche Computer Program (Appendix)</div> <div>6. <input type="checkbox"/> Nucleotide and/or Amino Acid Sequence Submission (if applicable, all necessary)<ul style="list-style-type: none">a. <input type="checkbox"/> Computer Readable Copyb. <input type="checkbox"/> Paper Copy (identical to computer copy)c. <input type="checkbox"/> Statement verifying identity of above copies</div>			
		ACCOMPANYING APPLICATION PARTS			
		<div>7. <input checked="" type="checkbox"/> Assignment Papers (cover sheet & document(s))</div> <div>8. <input type="checkbox"/> 37 CFR 3.73(b) Statement <input type="checkbox"/> Power of Attorney</div> <div>9. <input type="checkbox"/> English Translation Document (if applicable)</div> <div>10. <input type="checkbox"/> Information Disclosure <input type="checkbox"/> Copies of IDS Statement (IDS)/PTO-1449 Citations</div> <div>11. <input type="checkbox"/> Preliminary Amendment</div> <div>12. <input checked="" type="checkbox"/> Return Receipt Postcard (MPEP 503) (Should be specifically itemized)</div> <div>13. <input type="checkbox"/> Certified Copy of Priority Document(s) (if foreign priority is claimed)</div> <div>14. <input type="checkbox"/> Other :</div>			
15. If a CONTINUING APPLICATION, check appropriate box and supply the requisite information below and in a preliminary amendment: <input type="checkbox"/> Continuation <input type="checkbox"/> Divisional <input type="checkbox"/> Continuation-in-part (CIP) of prior Application No: Prior application information: Examiner: Group/Art Unit: For CONTINUATION or DIVISIONAL APPS only: The entire disclosure of the prior application, from which an oath or declaration is supplied under Box 4b, is considered a part of the disclosure of the accompanying continuation or divisional application and is hereby incorporated by reference. The incorporation can only be relied upon when a portion has been inadvertently omitted from the submitted application parts.					
16. CORRESPONDENCE ADDRESS					
<input type="checkbox"/> Customer Number or Bar Code Label (Insert Customer No. or Attach bar code label here) or <input checked="" type="checkbox"/> Correspondence address below					
NAME	Samuel H. Dworesky				
ADDRESS	AT&T CORP. P.O. Box 4110				
CITY	Middletown	STATE	New Jersey	ZIP CODE 07748-4110	
COUNTRY	United States of America			FAX 732-368-6932	
17. SIGNATURE OF APPLICANT, ATTORNEY, OR AGENT REQUIRED					
NAME	Susan E. McHale		Reg. #	35948	
TELEPHONE	908-221-5776				
SIGNATURE			DATE	11/07/2000	
"Express Mail" Mailing Label Number EL580483135US Date of Deposit 11/07/2000					
I hereby certify that this Patent Application is being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 CFR 1.10 on the date indicated above and is addressed to the Assistant Commissioner for Patents, Washington D.C. , 20231					
Dawn Cooper (Printed Name of Person Mailing Paper)					
 (Signature of Person Mailing Paper)					

FEE TRANSMITTAL

Patent Fees are subject to annual revision

Complete if Known

Application Number	
Filing Date	
First Named Inventor	Yuguang Wu
Examiner Name	
Group/Art Unit	
Attorney Docket No.	113122

TOTAL AMOUNT OF PAYMENT	\$830
-------------------------	-------

METHOD OF PAYMENT (check one)

1. The Commissioner is hereby authorized to charge indicated fees and credit any overpayments to:

Deposit Account Number	01-2745
Deposit Account Name	AT&T CORP.

☒ Charge Any Additional Fee Required Under 37 CFR 1.16 and 1.17
 ☐ Charge the Issue Fee Set in 37 CFR 1.18 at the Mailing Date of the Notice of Allowance
FEE CALCULATION**1. FILING FEE**

Large Fee Code	Entity Fee(\$)	Fee Description	Fee Paid
101	710	Utility(37CFR 1.53(b)) / CPA(37CFR 1.53(d)) Filing Fee	710
106	320	Design Filing Fee	
108	710	Reissue Filing Fee	
114	150	Provisional Filing Fee	

SUBTOTAL (1) 710**2. CLAIMS**
☐ Filing Under 37CFR 1.53 (b)
☐ CPA Under 37CFR 1.53 (d)
☐ Amendment
Extra Claims

Fee from below

Fee Paid

Total	12	- 20 =	0	X	18	=	0
Ind.	4	- 3 =	1	X	80	=	80
Multiple Dependent Claims						=	0

Large Fee Code	Entity Fee(\$)	Fee Description
103	18	Claims in excess of 20
102	80	Independent Claims in excess of 3
104	270	Multiple Dependent Claims
109	80	Reissue independent claims over original patent
110	18	Reissue claims in excess of 20 and over original patent

SUBTOTAL (2) 80**FEE CALCULATION (continued)****3. ADDITIONAL FEES**

Large Fee Code	Entity Fee(\$)	Fee Description	Fee Paid
105	130	Surcharge - late filing fee or oath	
127	50	Surcharge - late provisional filing fee or cover sheet	
139	130	Non-English specification	
147	2520	For filing a request for reexamination	
112*	920	Requesting publication of SIR prior to Examiner action	
113*	1840	Requesting publication of SIR after Examiner action	
115	110	Extension for reply within first month	
116	390	Extension for reply within second month	
117	890	Extension for reply within third month	
118	1390	Extension for reply within fourth month	
128	1890	Extension for reply within fifth month	
119	310	Notice of Appeal	
120	310	Filing a brief in support of an appeal	
121	270	Request for oral hearing	
138	1510	Petition to institute a public use proceeding	
140	110	Petition to revive - unavoidable	
141	1240	Petition to revive - unintentional	
142	1240	Utility issue fee (or reissue)	
143	440	Design issue fee	
122	130	Petitions to the Commissioner	
123	50	Petitions related to provisional applications	
126	240	Submission of Information Disclosure Statement	
581	40	Recording each patent assignment per property (times number of properties)	40
146	710	Filing a submission after final rejection(37 CFR 1.129(a))	
149	710	For each additional invention to be examined (37 CFR 1.129(b))	
Other fee (specify)			
Other fee (specify)			
* Reduced by Basic Filing Fee Paid			
SUBTOTAL(3)			40

SUBMITTED BY

Typed or Printed Name Susan E. Mchale

Complete (if applicable)

Reg. Number 35948

Signature

Susan E. Mchale

Date

11/7/00

Deposit Account User ID

SEND TO: Assistant Commissioner for Patents, Washington, D.C. 20231

JUS13 U.S. PTO 09/707474

11/07/00

A GENERIC FAULT-TOLERANT LOCKING MECHANISM
FOR MULTIPROCESSOR SYSTEMS

FIELD OF THE INVENTION

5 The present invention relates generally to a fault tolerant locking mechanism and to the use of such a mechanism in preventing deadlocks in multiprocessor computer systems.

BACKGROUND OF THE INVENTION

10 As used herein, the term computer includes any device or machine capable of accepting data, applying prescribed processes to the data, and supplying the results of the processes. A multiprocessing computer system has multiple processes executing on the system. Each process performs a particular task, and the processes, taken as a whole, perform some larger task, typically called an application. These processes may be executing on a single central computer or they may be running on separate computers which
15 are connected to each other via some type of communications link, i.e., a distributed or networked computer system.

 In multiprocessing systems, resources are often shared among the executing processes. Such resources may include, for example, disk drives, printers, shared memory and databases. During processing, a process may require exclusive access to a resource,
20 such that another process may not use that resource until the first process is finished with it. Thus, several processes may compete for a finite number of resources. This is commonly known as mutually exclusive sharing of resources.

 A problem with mutually exclusive access to resources is the possibility that the computer system will enter a deadlock state. A deadlock state is a state in the computer system in which, because of a resource allocation pattern, the computer system cannot progress past a processing point. For example, consider a computer system with two
25 resources R1 and R2, and two processors P1 and P2, where P1 and P2 both need simultaneous exclusive access to R1 and R2 at some point in order to successfully complete their processing. If P1 gains exclusive access to R1 and R2, then P2 must wait until P1
30 releases the resources. In this situation, P2 is described as being in a "wait" state. Such a situation does not present a problem, because, it is presumed, that P1 will eventually release the resources, at which time P2 may gain access to the resources. However, consider the situation in which P1 gains access to and holds R1, and P2 gains access to and holds R2, as

shown in FIG. 1. If this occurs, then P1 cannot finish its task until it gains access to R2, and P2 cannot finish its task until it gains access to R1, i.e., both P1 and P2 will enter a wait state. However, P1 is holding R1 and will not release R1 until it gains access to R2, and P2 is holding R2 and will not release R2 until it gains access to R1. At this point, the system is in a deadlock state.

Systems and mechanisms for deadlock avoidance and recovery are known in the art and described in U.S. Patents 5,664,088 to Romanovsky et al. and 5,913,060 to Discavage.

Generally, in the prior art, once a deadlock is detected, one of the processes involved in the deadlock is terminated, so that it releases the resource it held, and the resource can be reclaimed by the system. The reclaimed resource may then be used by a waiting process. If the waiting process can finish processing using the reclaimed resource, then the system can progress past the deadlock state. The terminated process, called the victim, is generally selected on a random basis, or based on a static priority assigned to the processes.

In addition to the "circular" deadlock situation shown in FIG. 1, a deadlock may also be caused by processors failing, or "crashing," before they release a resource. For example, in the same two processor system described above, assume processor P1 holds resource R1, and processor P2 holds resource R2 and is waiting for R1, as shown in FIG. 2. If processor P1 then unexpectedly crashes while still holding resource R1, P2 remains in a wait state, aware that R1 is busy, but not aware that P1 has crashed. Thus, P2 will wait forever for R1, which is being held by crashed processor P1, causing a deadlock state. Since prior art deadlock avoidance schemes are based on the assumption that processors will never malfunction, these schemes will not prevent deadlocks caused by crashing processors.

In any distributed or parallel computer system, access to shared resources is controlled by some form of a locking mechanism or scheme, whereby a shared resource is committed to, or "locked" by, a single holding processor until that processor releases the resource (i.e., releases the "lock"). To make such a system fault-tolerant, when a lock holding processor (i.e., a processor that has locked a resource) unexpectedly crashes, a rescuing method is needed to prevent system deadlock. The rescue operation will inherently be determined by the locking scheme. Problems may arise when two or more processors try to rescue the same locked resource. The correct behavior is that exactly one of them would succeed. Thus, there is a need for a generic fault-tolerant locking mechanism or scheme that can avoid deadlocks caused by failing processors in a multiprocessor system, especially mission-critical parallel systems where high availability is absolutely necessary. Such a

mechanism will enable waiting processors to identify a locked resource held by a failed processor, and "rescue" the resource from the hold of its failed processor, by changing, or "re-setting," a lock associated with the resource. A single waiting processor will then rescue a resource to prevent the potential deadlock from locking up all other processors waiting for
5 the same resource.

SUMMARY OF THE INVENTION

In a preferred embodiment, the present invention is a match-and-set lock for controlling access to a resource that is shared among a plurality of users N . The lock has a
10 locked operating state and an unlocked operating state controlled by a value C such that the lock is in its locked operating state when $C \neq 0$ and in its unlocked operating state where $C=0$. The lock returns a value R , equal to the lock's current content C , to an inquiring user seeking access to the resource. A return value $R=0$ usually denotes that the resource is free, and a return value $R \neq 0$ denotes that the resource is locked by another user. The lock is
15 responsive to an atomic command in the form (A, B) , such that the lock substitutes B for C if $A=C$. Thus, the lock may be obtained/locked by issuing the command (A, B) where $A=C$ and $B \neq 0$; and the lock may be released by issuing the command (A, B) where $A=C$ and $B=0$.

In accordance with the invention, a deadlock condition may be avoided by setting
20 the lock to the value $B=P+T*(N+1)$. Here N is the total number of users, P is an integer within $[1, N]$ that identifies the current user issuing the command (A, B) , and T is the current global time stamp. When the lock is set in this fashion, a return value of $R \neq 0$, identifies the user currently locking the resource (via $P'=R \bmod (N+1)$) and the time when that user locked the resource (via $T'=R/(N+1)$). If the inquiring user determines that the
25 user currently locking the resource has failed or restarted since locking the resource, this inquiring user can reset the lock by issuing the command (R, B) . Here $B=P+T*(N+1)$, P identifies this new user issuing the command, and T is the current time stamp; in so doing, the inquiring user gains access to the resource. If there are multiple users trying to issue this type of command (R, B) (of course, with different values of B for each inquiring user),
30 then exactly one of them will succeed. This is because the very first of them that has succeeded will have changed the content of the lock to something different from R , causing all others to fail. When the inquiring user is finished with the resource, the user can reset the current content of the lock to $C = 0$, to signal to other users that the resource is again

free. The plurality of users may be processors in a multiprocessor computer system, and the processors may form a telecommunications system.

In another preferred embodiment, the invention is a software program implemented in a multiprocessor computer system, the software program controlling access to a shared resource from a plurality of processors N, by employing a match-and-set lock that has a current content C and that is responsive to a match-and-set command (A, B) that writes B into the current content if A=C. The lock commits the shared resource to the control of that processor that most recently set the current content to a non-zero value. The software program enables the multiprocessor system to: (a) initialize the lock by setting the current content to zero; (b) issue from one of the processors a command (M, S) where M equals an expected current content of the lock and $S=P+T*(N+1)$ where $P < (N+1)$ and identifies the processor issuing the command and T is a current time stamp; (c) if a non-zero value R is returned from the lock, identify the processor that set the lock as $P'=R\% (N+1)$ where $R\% (N+1)$ is the remainder of R, mod (N+1), and identify the approximate locking time as $T'=R/(N+1)$; (d) if processor P' is still operating and was last started before time T', repeat step (b); in the alternative, (e) if processor P' is not still operating or was last started after time T', issue from processor P the command (M, S) where $M=R$. The users may be processors in a multiprocessor computer system, and the processors may form a telecommunications system.

In still another preferred embodiment, the invention is a method for controlling access to a shared resource from a plurality of processors N, comprising the steps of: (a) providing the resource with a locking storage function, wherein the function stores either an initial value denoting that the resource is not currently held by a processor or a value denoting a holding processor, currently holding the resource, if any, and an initial holding time when such holding began; (b) initializing the function by storing an initial value denoting that the resource is not currently held by a processor; (c) querying the function to determine if the resource is currently held by a processor; (d) if the resource is not currently held by a processor, providing to an inquiring processor access to the resource and resetting the value of the function to denote the inquiring processor's identity and a current time; in the alternative, (e) if the resource is currently held by a processor, determining if the holding processor has failed or been restarted since taking hold of the resource; (f) if the holding processor has failed or been restarted since taking hold of the resource, providing to the inquiring processor access to the resource and resetting the value of the function to denote the inquiring processor's identity and a current time; in the alternative, (g) if the

holding processor has not failed or been restarted since taking hold of the resource, repeating steps (b) through (d); and (h) upon release of the resource, resetting the value stored in the function to an initial value denoting that the resource is not currently held by a processor. The plurality of processors N may form a telecommunications system.

5 In still another preferred embodiment, the invention is a method of controlling access to a shared resource from a plurality of processors N, the method employing a match-and-set lock that has a current content C and that is responsive to a match-and-set command (A, B) that writes B into the current content if A=C. The lock commits the shared resource to the control of that processor that most recently set the current content to
 10 a non-zero value. The method comprises the steps of: (a) initializing the lock by setting the current content to zero; (b) issuing from one of the processors a command (M, S) where M equals an expected current content of the lock and $S=P+T*(N+1)$ where $P < (N+1)$ and identifies the processor issuing the command and T is a current time stamp; (c) if a non-zero value R is returned from the time lock, identifying the processor that set the lock as $P'=R\%$
 15 $(N+1)$ where $R\%(N+1)$ is the remainder of R, mod (N+1), and identifying the approximate locking time as $T'=R/(N+1)$; (d) if processor P' is still operating and was last started before time T', repeating step (b); in the alternative, (e) if processor P' is not still operating or was last started after time T', issuing from processor P the command (M, S) where $M=R$. The plurality of processors N may form a telecommunications system.

20 Using pseudo algorithmic language, the invention of fault-tolerant locking mechanism can be summarized as follows: here time() returns the current global time, N is the total number of users, my_id() returns the calling user's id (between 1 and N), alive(p) returns TRUE is user p is still running, and start_time(p) returns the latest reboot time of user p.

25 Lock:
 M = 0;
 S = my_id() + time() * (N + 1);
 While ((R = **match-and-set**(M, S)) != M) {
 30 Who = R%(N + 1);
 When = R/(N + 1);
 If (!alive(who) || start_time(who) > when)
 M = R;
 Else
 35 M = 0;
 S = my_id() + time() * (N + 1);
 Unlock:

match-and-set(0, S).

BRIEF DESCRIPTION OF THE DRAWINGS

The present invention will be understood and appreciated more fully from the following detailed description, taken in conjunction with the drawings in which:

FIG. 1 is a block diagram illustrating a multiprocessor system in a typical deadlock state.

FIG. 2 is a block diagram illustrating a multiprocessor system in a deadlock state caused by a failed processor.

FIGS. 3-7 are block diagrams illustrating the general operation of a match-and-set lock in a preferred embodiment of the present invention.

FIG. 8 is a flowchart illustrating the steps in a preferred embodiment of the method of the present invention.

FIG. 9 is an equivalent flowchart illustrating the steps of the same preferred embodiment of the method of the present invention.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

Reference is now made to FIGS. 3 through 7, which are block diagrams illustrating the general operation of a match-and-set lock in a preferred embodiment of the present invention. The lock has a locked operating state and an unlocked operating state controlled by a value C such that the lock is in its locked operating state when $C \neq 0$ and in its unlocked operating state where $C=0$. The lock returns a value R, equal to the lock's current content C, to an inquiring user seeking access to the resource. A return value $R=0$ denotes that the resource is free, and a return value $R \neq 0$ denotes that the resource is locked by another user.

When the return value $R \neq 0$, R identifies the user currently locking the resource and the time when that user first locked the resource. More particularly, in accordance with the invention, $R=P+T*(N+1)$ where N is the number of users, $P < (N+1)$ is the identifier of the user who locked the resource and T is the current time stamp at the time the resource was locked. General requirements of the environment in which the match-and-set lock is applied are that: there is a global reference time, and any user can determine whether any other user is currently running or has crashed, as well as the time of each user's latest reboot time. Note that all these data are easy to implement: every user maintains a reboot time stamp; upon coming online, the user would read from the global time clock and save in its time stamp.

In FIG. 3, a shared resource 30 has a match-and-set lock 32, which has a current content R. The lock is responsive to a match-and-set command (M, S) that writes S into the current content if $M=R$, and commits the shared resource 30 to the control of the processor that most recently set the current content of the lock to a non-zero value. In FIG. 3, processor 34 has gained access to the resource, and the current content R of lock 32 denotes that processor 34 has locked resource 30 at a time T' . However, processor 34 has unexpectedly crashed while still holding resource 30. Some time later, inquiring processor 36, in an attempt to gain access to resource 30, issues a match-and-set command (M, S) to lock 32, where $M=0$, and S denotes the identity of processor 36 and the current time.

As shown in FIG. 4, the lock 32 returns the current value, R, to the inquiring processor. Since $R \neq 0$, the inquiring processor 36 determines the identity of the holding processor, $P' = R \bmod (N+1)$, where $R \bmod (N+1)$ is the remainder of R, mod(N+1), and the time, $T' = R / (N+1)$, when the processor P' took control of the resource. Therefore, inquiring processor 36 can identify processor 34 as the holding processor and can determine if processor 34 has crashed or been rebooted since time T' . Since processor 34 has crashed, in FIG. 5, processor 36 issues a match-and-set command (M, S), where $M=R$, to lock 32. Since $M=R$, processor 36 will gain access to resource 30, effectively rescuing resource 30 from the "dead-hand" control of processor 34. If, however, processor 34 had not crashed or rebooted since time T' , inquiring processor 36 would continue to issue the command (M, S), where $M=0$ and $S=P+T*(N+1)$ with T being the up-to-date current time, to lock 32 until resource 30 became free, or until it determined that holding processor 36 had crashed or rebooted since time T' .

The value S, which denotes the identity of processor 36 and the time when processor 36 took control of resource 30, is now stored in lock 32, as shown in FIG. 6. At this time, an inquiring processor can determine the identity (i.e., a processor number or code assigned to processor 36) of holding processor 36 from the value S, the identity being equal to $S \bmod (N+1)$, which is the remainder of S, mod(N+1), and the approximate time when processor 36 took control of resource 30, which is equal to $S / (N+1)$. When processor 36 has completed its use of resource 30, it resets the value of lock 32 to zero, and resource 30 will be available for use by other potential inquiring processors 38, as shown in FIG. 7. Thus, the match-and-set lock of the present invention allows processors to rescue a resource from a "dead" processor. Moreover, because more than one processor can try to rescue the resource at the same time, it is important to note that the present invention ensures that only one such processor will succeed. The present invention is also extremely valuable in

mission-critical parallel systems, such as fault tolerant telecommunications systems, where high-availability is absolutely necessary, and the failure of individual processors should not bring down the entire system.

Reference is now made to FIG. 8, which is a flowchart illustrating the steps in a preferred embodiment of the method of the present invention. In step 80, a "match and set" lock that controls access to a shared resource is initialized by setting the current content C of the lock equal to zero. The resource is shared by a plurality of N processors. The lock is responsive to a match-and-set command (A, B), issued by a processor, that writes B into the current content of the lock if $A=C$, and commits the shared resource to the control of that processor that most recently set the current content of the lock to a non-zero value. In step 82, a processor seeking access to the shared resource issues the command (M, S), where $M=0$ equals the expected content of a free lock, and $S=P + T*(N+1)$ where $P < (N+1)$ and identifies the processor issuing the command and T is a current time stamp. In step 84, the lock returns a value R in response to the inquiring processor's query. In step 86, if $R=M=0$, the resource has been free, and the inquiring processor succeeds in locking the lock and gaining control over the resource in step 94. In this case, the lock content will have been set to the value of S in step 84. Here S denotes the identity of the processor currently holding the lock (i.e., an identification number or code assigned to the processor in question), and the time that the holding processor gained control, as described above. In step 96, when the holding processor has completed its use of the resource, it releases the resource and resets the lock content equal to zero, signaling to other inquiring processors that the resource is free.

Returning to step 86, if $R \neq 0$, the resource is currently held by another processor (i.e., another processor has already set the lock to a non-zero value). In step 88, the inquiring processor then determines the identity of the processor that set the lock, $P' = R \%(N+1)$ where $R \%(N+1)$ is the remainder of R, i.e. $\text{mod}(N+1)$, and the time, $T' = R/(N+1)$, when the lock was set by processor P' . In step 90, a test is made to determine if the holding processor P' has crashed or been restarted since time T' (the time when P' locked the resource). If so, the inquiring processor rescues the resource in step 92 by issuing the match-and-set command (M, S), with $M=R$. In step 98, the match-and-set command returns a new value R' in its response. If there is only one such processor trying to rescue this lock, then since the current content of the lock is equal to R, the match-and-set will succeed: namely $R'=M$ in step 99. In this case, the lock is set to the value S, and the inquiring processor gains control of the resource, in step 94. Again, in step 96, when the

holding processor has completed its use of the resource, it releases the resource and resets the lock content equal to zero, signaling to other inquiring processors that the resource is free. If there are more than one such processors trying to rescue this lock at the same time, only one of them will succeed: the others will all fail the test in step 99, which in turn go
5 back to step 82 and try again from the beginning.

Returning to the test in step 90, if the holding processor P' has not crashed, or been restarted, since time T', the inquiring processor returns to step 82, and repeats the subsequent steps, as necessary, until the resource is released by its holding processor, or the inquiring processor determines that the holding processor has crashed. In the present
10 method, a resource locked by a crashed processor in a multiprocessor system is rescued by the first processor to reset the lock with its processor ID and time stamp information, thereby eliminating the problems, and potential deadlocks, caused by multiple processors attempting to rescue the same resource at the same time.

An equivalent embodiment of the present invention is depicted in the flowchart of
15 FIG. 9. In step 100, a "match and set" lock that controls access to a shared resource is initialized by setting the current content M of the lock equal to zero in step 102. The resource is shared by a plurality of N processors. The lock is responsive to a match-and-set command (A, B), issued by a processor, that writes B into the current content of the lock if $A=C$, and commits the shared resource to the control of that processor that most recently set
20 the current content of the lock to a non-zero value. In step 102, a processor seeking access to the shared resource sets $M=0$ which is the expected content of a free lock. In step 103 it issues the command (M, S), and $S=P + T*(N+1)$ where $P < (N+1)$ and identifies the processor issuing the command and T is a current time stamp. In step 104, the lock returns a value R in response to the inquiring processor's query. In step 106, if $R=M$, the resource
25 has been free, and the inquiring processor succeeds in locking the lock and gaining control over the resource in step 116. Also in step 116, the lock content has been set to a value S, which denotes the identity of the processor currently holding the lock (i.e., an identification number or code assigned to the processor in question), and the time that the holding processor gained control, as described above. In step 118, when the holding processor has
30 completed its use of the resource, it releases the resource and resets the lock content equal to zero, signaling to other inquiring processors that the resource is free.

Returning to step 106, if $R \neq M$, the resource is currently held by another processor (i.e., another processor has already set the lock to a non-zero value). In step 108, the inquiring processor then determines the identity of the processor that set the lock, $P' =$

$R\%(N+1)$ where $R\%(N+1)$ is the remainder of R , i.e. $\text{mod}(N+1)$, and the time, $T' = R/(N+1)$, when the lock was set by processor P' . In step 110, a test is made to determine if the holding processor P' has crashed or been restarted since time T' (the time when P' locked the resource). If so, the inquiring processor sets $M=R$ in step 114, and goes back to step 103 to try to rescue the lock, by issuing the match-and-set command (M, S) , with $M=R$.

Returning to the test in step 110: if the holding processor P' has not crashed, or been restarted, since time T' , the inquiring processor sets $M=0$, the expected content of a free lock, and returns to step 103 and repeats the subsequent steps, as necessary, until the resource is released by its holding processor, or the inquiring processor determines that the holding processor has crashed.

While the present invention has been described with reference to the preferred embodiments, those skilled in the art will recognize that numerous variations and modifications may be made without departing from the scope of the present invention.

Accordingly, it should be clearly understood that the embodiments of the invention described above are not intended as limitations on the scope of the invention, which is defined only by the following claims.

WHAT IS CLAIMED IS:

- 1 1. A match-and-set lock for controlling access to a resource that is shared
2 among a plurality of users N , said lock having a locked operating state and an unlocked
3 operating state controlled by a value C such that the lock is in its locked operating state
4 when $C \neq 0$ and in its unlocked operating state where $C=0$, said lock being responsive to a
5 command in the form (A, B) , said lock substituting B for C if $A=C$, whereupon:
6 the lock may be released by a user issuing the command (A, B) where $A=C$
7 and $B=0$;
8 the lock may be set by a user issuing the command (A, B) where $A=C$ and
9 $B \neq 0$; and
10 a deadlock condition may be avoided by always setting the lock to the value
11 $B=P+T*(N+1)$ where $P < (N+1)$ and identifies the user issuing the command (A, B) and T is
12 the current time stamp.
- 1 2. The match-and-set lock of claim 1, wherein the lock returns a value R , equal
2 to the lock's current content C , to an inquiring user seeking access to the resource.
- 1 3. The match-and-set lock of claim 2, wherein a return value $R=0$ denotes that
2 the resource is free, and a return value $R \neq 0$ denotes that the resource is locked by another
3 user.
- 1 4. The match-and-set lock of claim 3, wherein the return value R , when $R \neq 0$,
2 identifies the user currently locking the resource and the time when that user first locked the
3 resource.
- 1 5. The match-and-set lock of claim 4, wherein, if the inquiring user determines
2 that the user currently locking the resource has failed or restarted since locking the resource,
3 the inquiring user can gain access to the resource and reset the lock by issuing the command
4 (R, D) , where $D=P+T*(N+1)$, $P < (N+1)$ identifies the current user issuing the command and
5 T is the current time stamp.
- 1 6. The match-and-set lock of claim 1, wherein the plurality of users N are
2 processors in a multiprocessor computer system.
- 1 7. The match-and-set lock of claim 6, wherein the processors form a
2 telecommunications system.
- 1 8. A software program implemented in a multiprocessor computer system, said
2 software program controlling access to a shared resource from a plurality of processors N ,
3 by employing a match-and-set lock that has a current content C and that is responsive to a
4 match-and-set command (A, B) that writes B into the current content if $A=C$, said lock

5 committing the shared resource to the control of that processor that most recently set the
6 current content to a non-zero value, said software program allowing the multiprocessor
7 system to :

- 8 (a) initialize the lock by setting the current content to zero;
- 9 (b) issue from one of the processors a command (M, S) where M equals
10 an expected current content of the lock and $S = P + T * (N + 1)$ where $P < (N + 1)$ and identifies
11 the processor issuing the command and T is a current time stamp;
- 12 (c) if a non-zero value R is returned from the time lock, identify the
13 processor that set the lock as $P' = R \% (N + 1)$ where $R \% (N + 1)$ is the remainder of R, mod
14 (N+1), and identify the approximate locking time as $T' = R / (N + 1)$;
- 15 (d) if processor P' is still operating and was last started before time T',
16 repeat step (b);
- 17 (e) if processor P' is not still operating or was last started after time T',
18 issue from processor P the command (M, S) where $M = R$.

1 9. A method of controlling access to a shared resource from a plurality of
2 processors N, comprising the steps of:

- 3 (a) providing the resource with a locking storage function, wherein the function
4 stores either an initial value denoting that the resource is not currently held by a processor
5 or a value denoting a holding processor currently holding the resource and an initial holding
6 time when such holding began;
- 7 (b) initializing the function by storing an initial value denoting that the resource
8 is not currently held by a processor;
- 9 (c) querying the function to determine if the resource is currently held by a
10 processor;
- 11 (d) if the resource is not currently held by a processor, providing to an inquiring
12 processor access to the resource and resetting the value of the function to denote the
13 inquiring processor's identity and a current time;
- 14 (e) if the resource is currently held by a processor, determining if the holding
15 processor has failed or been restarted since taking hold of the resource;
- 16 (f) if the holding processor has failed or been restarted since taking hold of the
17 resource, providing to the inquiring processor access to the resource and resetting the value
18 of the function to denote the inquiring processor's identity and a current time;
- 19 (g) if the holding processor has not failed or been restarted since taking hold of
20 the resource, repeating steps (b) through (d); and

21 (h) upon release of the resource, resetting the value stored in the function to an
22 initial value denoting that the resource is not currently held by a processor.

1 10. The method of claim 9 wherein the plurality of processors N form a
2 telecommunications system.

1 11. A method of controlling access to a shared resource from a plurality of
2 processors N, said method employing a match-and-set lock that has a current content C and
3 that is responsive to a match-and-set command (A, B) that writes B into the current content
4 if A=C, said lock committing the shared resource to the control of that processor that most
5 recently set the current content to a non-zero value, said method comprising the steps of:

6 (a) initializing the lock by setting the current content to zero;

7 (b) issuing from one of the processors a command (M, S) where M
8 equals an expected current content of the lock and $S=P+T*(N+1)$ where $P < (N+1)$ and
9 identifies the processor issuing the command and T is a current time stamp;

10 (c) if a non-zero value R is returned from the time lock, identifying the
11 processor that set the lock as $P'=R\% (N+1)$ where $R\% (N+1)$ is the remainder of R, mod
12 (N+1), and identifying the approximate locking time as $T'=R/(N+1)$;

13 (d) if processor P' is still operating and was last started before time T',
14 repeating step (b);

15 (e) if processor P' is not still operating or was last started after time T',
16 issuing from processor P the command (M, S) where $M=R$.

1 12. The method of claim 11 wherein the plurality of processors N form a
2 telecommunications system.

3

3

Abstract of the Disclosure

A generic fault tolerant "match-and-set" locking mechanism and method for controlling access to resources shared among a plurality of users N . The match-and-set lock has a locked operating state and an unlocked operating state controlled by the value C of its content such that the lock is in its locked operating state when $C \neq 0$ and in its unlocked operating state where $C=0$. The lock returns a value R , equal to the lock's current content C , to an inquiring user seeking access to the resource. A return value $R=0$ denotes that the resource is free, and a return value $R \neq 0$ denotes that the resource is locked by another user. The lock is responsive to a command in the form (A, B) , such that the lock substitutes B for C if $A=C$. Thus, the lock may be locked by issuing the command (A, B) where $A=C$ and $B \neq 0$; and the lock may be released by issuing the command (A, B) where $A=C$ and $B=0$. A deadlock condition may be avoided by setting the lock to the value $B=P+T*(N+1)$, where $P < (N+1)$ and identifies the user issuing this command (A, B) , and T is the current global time stamp when the user issues this command. When the return value $R \neq 0$, R identifies the user currently locking the resource and the time when that user locked the resource: the locking user id is simply $R\%(N+1)$, and the locking time is $R/(N+1)$. An inquiring user can then determine if the user currently locking the resource has failed, or restarted, since locking the resource. If so, the inquiring user can rescue the lock from the control of the failed user. The mechanism of "match-and-set" ensures that exactly one such user can succeed in rescuing that lock. Crashed users can come back online and participate in new locking operations: the locking semantics is always maintained and this system suffers no deadlocks. The implementation requirements are easy to satisfy: that there is a global reference clock, that any user maintains a time stamp (off the global clock) of its latest reboot time, and the match-and-set access is an indivisible operation. The last can be implemented in hardware in a way similar to that of the traditional test-and-set lock. The plurality of users may be processors in a multiprocessor computer system, and the processors may form a telecommunications system.

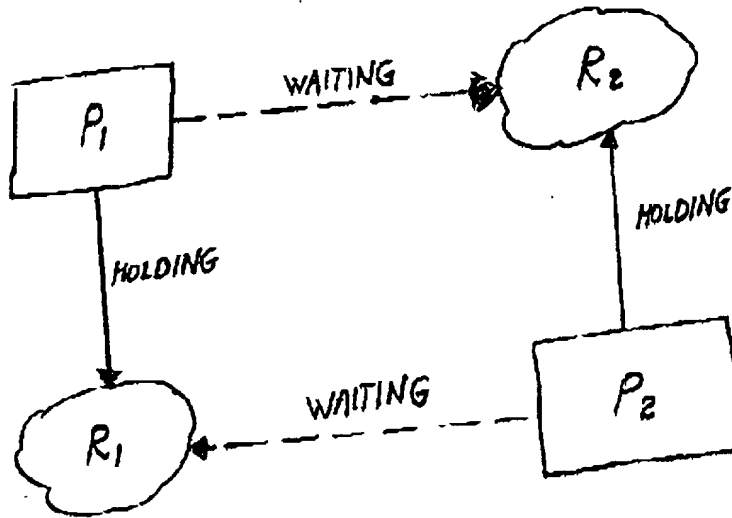


FIG. 1

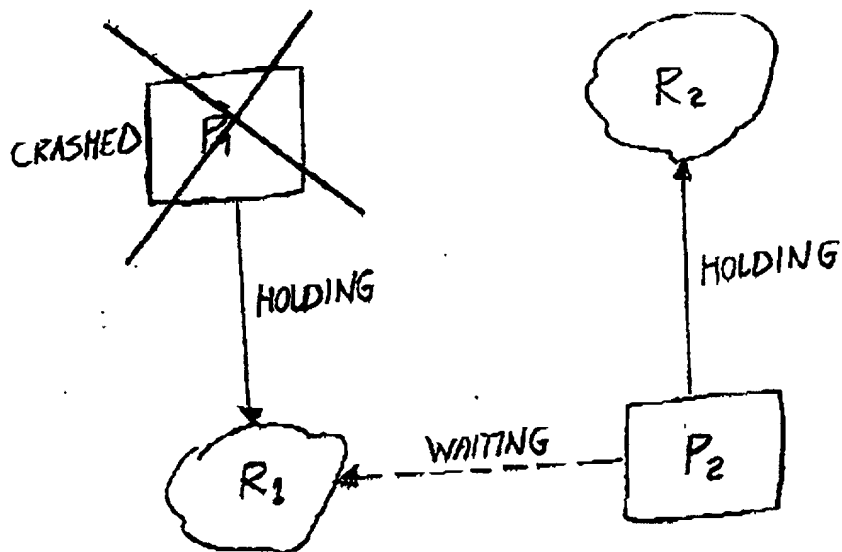


FIG. 2

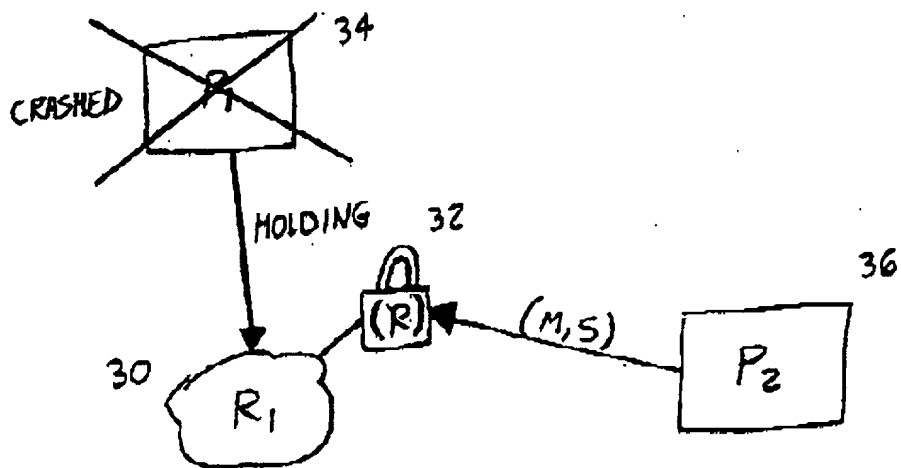


FIG. 3

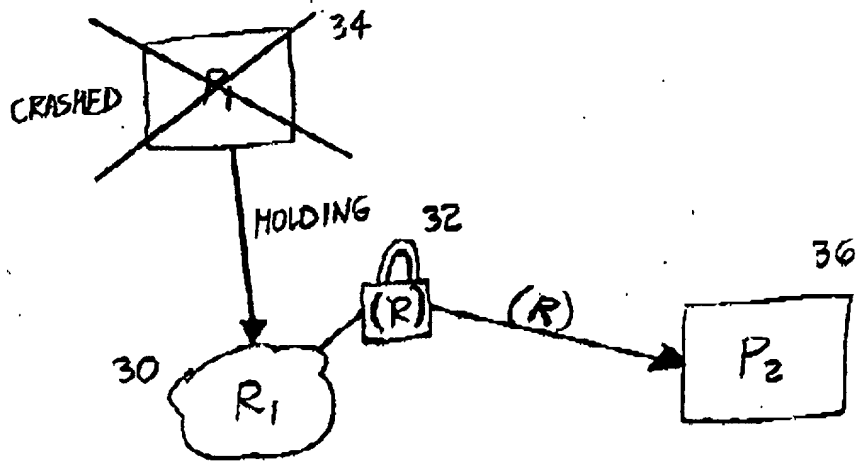
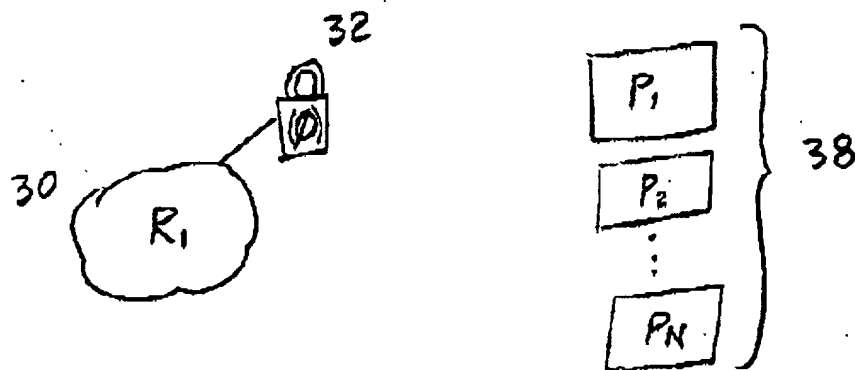
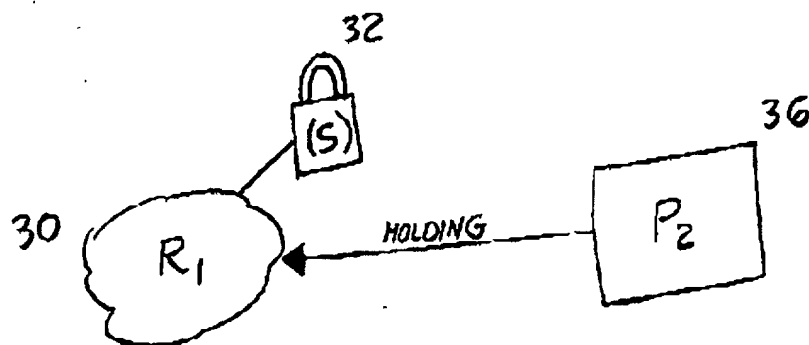
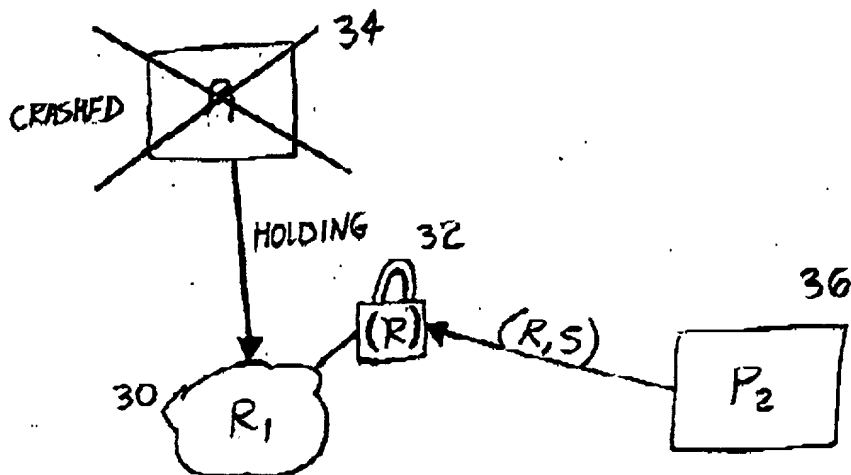
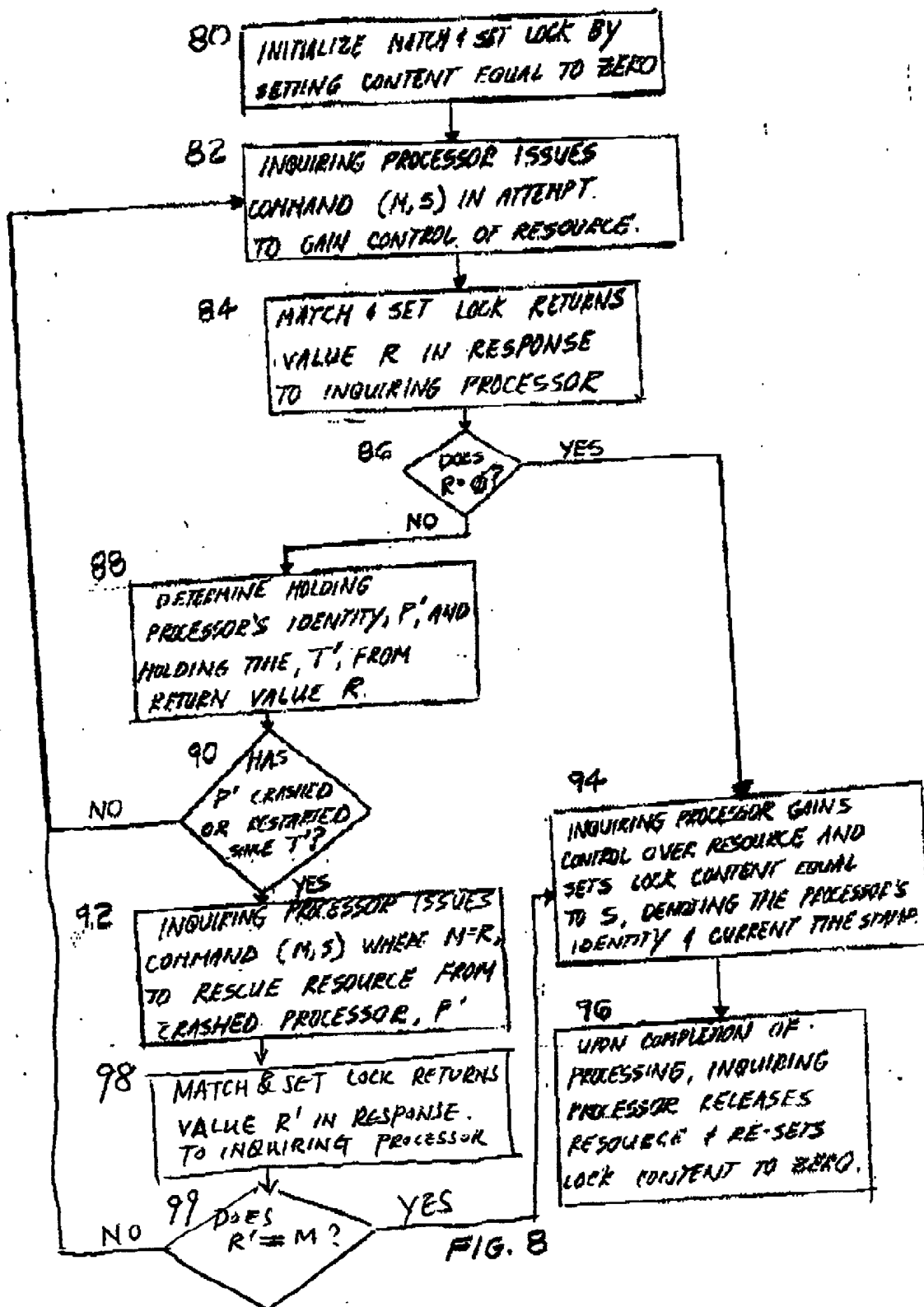


FIG. 4





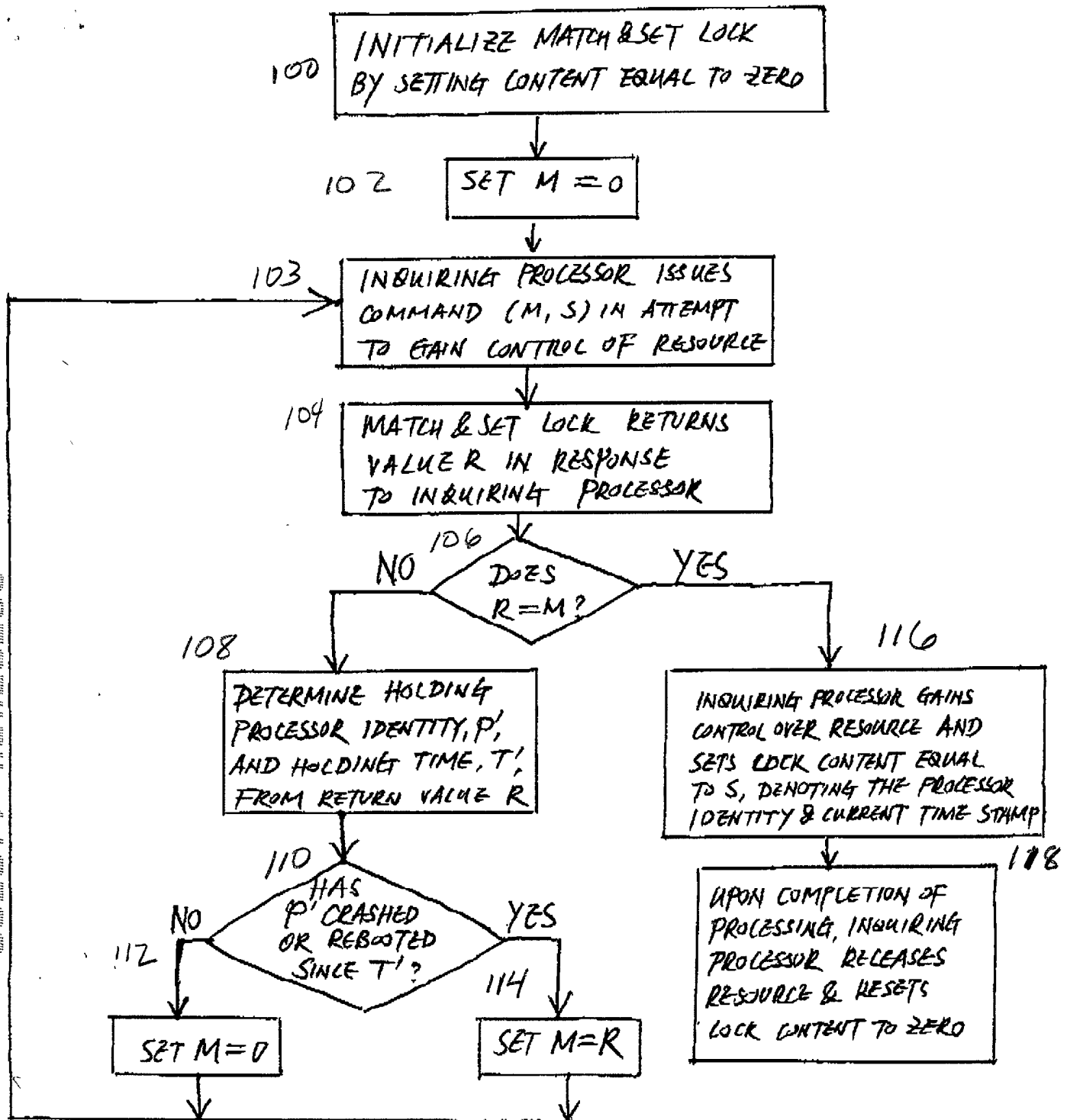


FIG. 9

DECLARATION FOR UTILITY OR DESIGN PATENT APPLICATION <input checked="" type="checkbox"/> Declaration Submitted with Initial Filing <input type="checkbox"/> Declaration submitted after Initial Filing	Attorney Docket Number	113122
	First Named Inventor	Yuguang Wu
	COMPLETE IF KNOWN	
	Application Number	
	Filing Date	
	Group Art Unit	
	Examiner Name	

As a below named inventor, I hereby declare that:

My residence, post office address, and citizenship are as stated below next to my name.

I believe I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if plural names are listed below) of the subject matter which is claimed and for which a patent is sought on the invention entitled:

A Generic Fault-Tolerant Locking Mechanism for Multiprocessor Systems

(Title of Invention)

the specification of which

☒ is attached hereto

OR

☐ was filed on

as United States Application Number or PCT International

Application Number

and was amended on

(if applicable).

I hereby state that I have reviewed and understand the contents of the above identified specification, including the claims, as amended by any amendment specifically referred to above.

I acknowledge the duty to disclose information which is material to patentability as defined in Title 37 Code of Federal Regulations, § 1.56.

I hereby claim foreign priority benefits under Title 35, United States Code § 119 (a)-(d) or § 365(b) of any foreign application(s) for patent or inventor's certificate, or § 365(a) of any PCT international application which designated at least one country other than the United States of America, listed below and have also identified below, by checking the box, any foreign application for patent or inventor's certificate, or of any PCT international application having a filing date before that of an application on which priority is claimed.

Prior Foreign Application Number(s)	Country	Foreign Filing Date (MM/DD/YYYY)	Priority Not Claimed	Certified Copy Attached?	
			<input type="checkbox"/>	YES	NO
			<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
			<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
			<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

☐ Additional foreign application numbers are listed on a supplemental priority data sheet PTO/SB/02B attached hereto

I hereby claim the benefit under 35 U.S.C. 119(e) of any United States provisional application(s) below.

Application Number(s)	Filing Date(MM/DD/YYYY)	
		<input type="checkbox"/> Additional provisional application numbers are listed on a supplemental priority data sheet PTO/SB/02B attached hereto

SEND TO: Assistant Commissioner for Patents, Box Patent Application, Washington, D.C. 20231

Attorney Docket Number: 113122

DECLARATION - Utility or Design Patent Application

I hereby claim the benefit under 35 U.S.C. 120 of any United States application(s), or 385(e) of any PCT international application designating the United States of America, listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in the prior United States or PCT International application in the manner provided by the first paragraph of 35 U.S.C. 112, I acknowledge the duty to disclose information which is material to patentability as defined in 37 C.F.R. 1.56 which became available between the filing date of the prior application and the national or PCT International filing date of this application.

U.S. Parent Application or PCT Parent Number	Parent Filing Date (MM/DD/YYYY)	Parent Patent Number (If applicable)

☐ Additional U.S. or PCT International application numbers are listed on a supplemental priority data sheet PTO/SB/02B attached hereto.

As a named inventor, I hereby appoint the following registered practitioner(s) with full power of substitution and revocation, to prosecute this application, to make alterations and amendments therein, to receive the patent, and to transact all business in the Patent and Trademark Office connected therewith:

☐ Customer Number

Place Customer Number Bar Code Label here

OR

☒ Registered practitioner(s) name/registration number listed below

Name	Registration Number	Name	Registration Number
CONOVER, Michele L.	34962	DELACRUZ, Cedric G	36498
DWORETSKY, Samuel H.	27873	GARG, Rohini K	45272
ISAACSON, Thomas M.	44166	LEE, Benjamin S.	42787
LEVY, Robert B.	28234	MCHALE, SUSAN E.	35948
MONKA, Gary H.	35290	NAVON, Jeffrey M	32711

☒ I also appoint the following additional registered practitioner(s) named on the Registered Practitioner Information (Supplemental Sheet) (PTO/SB/02C modified by AT&T Corp.) attached hereto with full power of substitution and revocation, to prosecute this application, to make alterations and amendments therein, to receive the patent, and to transact all business in the Patent and Trademark Office connected therewith.

Direct all Correspondence to:

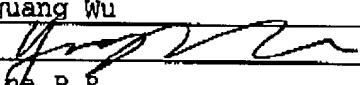
☐ Customer Number or Bar Code Label

(Insert Customer No. or Attach bar code label here)

or ☒ Correspondence address below

NAME	Samuel H. Dworetsky				
ADDRESS	AT&T CORP. P.O. Box 4110				
CITY	Middletown	STATE	New Jersey	ZIP CODE	07748-4110
COUNTRY	United States of America			FAX	732-368-6932

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under 18 U.S.C. 1001 and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

Name of Sole or First Inventor		<input type="checkbox"/> A petition has been filed for this unsigned inventor	
Name	Yuguang Wu		
Signature			Date
Citizenship	China P.R.		
Address (line 1)	350 Wedgewood Road		
Address (line 2)	Morganville		
Address (line 3)	Monmouth County		
Address (line 4)	New Jersey		
Address (line 5)	USA		
Zip Code	07751		

☐ Additional Inventors are being named on the separately numbered sheets attached hereto

SEND TO: Assistant Commissioner for Patents, Box Patent Application, Washington, DC 20231

Attorney Docket Number: 113122

DECLARATION -- Supplemental Priority Data Sheet

Additional Foreign Applications:

[illegible]

Additional Provisional applications:

Application Number(s)	Filing Date(MM/DD/YYYY)

Additional U.S. applications:

U.S. Parent Application or PCT Parent Number	Parent Filing Date (MM/DD/YYYY)	Parent Patent Number (if applicable)

SEND TO: Assistant Commissioner for Patents, Box Patent Application, Washington, DC 20231

Attorney Docket Number: 113122

DECLARATION		Registered Practitioner Information (Supplemental Sheet)	
Name	Registration Number	Name	Registration Number
RESTAINO, Thomas A.	33444	STEINMETZ, Alfred G.	22971